DOCUMENT RESUME

ED 090 247 TM 003 509

AUTHOR Tuckman, Bruce W.
TITLE The Differentiated Outcome Hypothesis or, When Will We Stop Using Conventional Achievement as the Sole Criterion for Evaluating Unconventional Instruction.
PUB DATE Apr 74
NOTE 11p.; Paper presented at the American Educational Research Association Annual Meeting (Chicago, Illinois, April 15-19, 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Academic Achievement; Attitudes; Behavior Change; *Evaluation Criteria; *Evaluation Needs; *Experimental Programs; Measurement; Models; Problem Solving; *Program Evaluation; *Summative Evaluation; Test Validity; Values

ABSTRACT

The differentiated outcome hypothesis posits that the maximum effects of an educational program are realized along those dimensions emphasized in the program or in those central to the persons involved. The idea that educational programs employing unconventional means should be examined on outcomes that are consistent with those means is derived from this hypothesis. In addition to knowledge acquisition outcomes, thinking and problem-solving, attitude and value, and behavioral change outcomes are recommended. Studies are cited that utilized outcome measures from the above classes and successfully demonstrated the differentiated merits of the programs involved. Relevant studies in progress are also reviewed. (Author)

Session # 21.14

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 090247

AERA PAPER, DIVISION D
CHICAGO, ILLINOIS
APRIL, 1974

The Differentiated Outcome Hypothesis or When Will We Stop Using Conventional Achievement as the Sole Criterion for Evaluating Unconventional Instruction

Bruce W. Tuckman
Rutgers University

TM 003 509

1. The Problem

It stands to reason that different kinds of interventions and programs should result in different outcomes, yet the overriding tendency is to rely on standardized achievement data as the basic criteria for evaluation. When an "experimental" program fails to produce greater gains in standardized achievement than the "conventional" or "traditional" program, the inevitable conclusion drawn is that the experimental program is therefore not worth the extra trouble. Perhaps this attitude reflects a tendency by educators, parents, and, all too often, evaluators, to rely on a limited number of outcome measures - often restricted to the standardized achievement variety. This is defended on the grounds that (a) we get these data anyway (i.e., standardized achievement testing is a part of the regimen of most school districts); (b) other things cannot be measured as easily nor as well (i.e., how do you measure liking for school, creativity, etc.); (c) producing achievement of the standardized achievement-test variety is the nitty-gritty of education; all the other stuff - e.g., creativity, attitudes, self-discipline - are frills.

2. The Conception

The major issues for measurement specialists and evaluators to contribute to are twofold: first, what outcomes should you be measuring in your evaluation of a program or other form of intervention, and second,

how do you measure these outcomes. The idea that different programs when successful yield different outcomes peculiar to those programs I call the differentiated outcome hypothesis. It may also be applied to individual differences; i.e., different students will maximize performance on different outcomes as a function of their characteristic way of relating to their environment.

The differentiated outcomes hypothesis suggests that evaluators consider a range of outcomes that may accurately reflect the results of the treatment being evaluated. Three trends appear in many evaluations,* namely: (1) the input side is considerably differentiated, i.e., treatments are described in detail - often with respect to some instructional model, (2) the outcome side is often restricted to achievement measurement alone, (3) little concern is given to individual differences in students other than to control for them. Since achievement measurement has evolved alongside the basic classroom, group-oriented instructional procedure which we tend to label "conventional" or "traditional," it is not unreasonable to expect that the conventional instructional approach will maximize achievement. If this expectation is correct, then the use of achievement as a sole evaluation criterion would favor the conventional or tried-and-true approach to instruction. Such evaluation would have to be considered short-sighted and even potentially biased in its limitations.

The position to be offered in this paper, based on the differentiated outcome hypothesis, is that evaluation studies should include a range of outcome measures specifically chosen to sample outcomes that are different not only in degree but in kind.

---

*It must be emphasized that the majority of educational evaluations are unpublished. Published instructional research does not fall into this threefold pattern to the same degree as do evaluation studies.

3. Some Support

Gagné (1971, 1972) has distinguished five domains for classifying learning processes, namely: (1) verbal information, (2) intellectual skills, (3) cognitive strategies, (4) motor skills, and (5) attitudes. These categories can be used as well for classifying outcomes and suggest a range of areas in which measurement of program effects might be undertaken. Glaser (1972 p. 12) also suggests measurement outside the "narrow band of traditional academic outcomes," with the inclusion of "measures of process and style, of cognitive and non-cognitive development, and of performance in more natural settings." He also argues strongly in favor of the inclusion of individual difference measures of cognitive process within instructional research. Lohnes (1973) while arguing for the inclusion of intellectual development as an important outcome of many (possibly all) instructional treatments, cites arguements for the inclusion of differentiated outcomes. Tyler (1951) said quite succinctly:

> "Any learning situation has multiple outcomes. While the child is acquiring information, knowledges, and skills, there is also taking place concomitant learnings in attitudes, appreciations, and interests. This view indicates a shift from a narrow conception of subject-matter outcomes to a broader conception of growth and development of individuals." (p. 48).

And finally, Cronbach (1971, p. 460) states: "Consequently, an ideally suitable battery for evaluation purposes will include separate measures of all outcomes the users of the information consider important."

4. The Model

I would like to suggest five categories or classes of outcomes, based on pronouncements like those above, that might reasonably form the basis for the application of the differentiated outcome hypothesis and which are represented in Figure 1. These are: (a) specific knowledge and

A. SPECIFIC KNOWLEDGE AND COMPREHENSION (SPEC COG)
   Subject-matter achievement (e.g., knowledge and understanding of math, science, language arts, social studies, machine operation, etc., based on specific curriculum content and specific objectives of program.)

B. GENERAL KNOWLEDGE AND COMPREHENSION (GEN COG)
   intelligence
   aptitudes
   general abilities
   reading ability
   reasoning ability
   knowledge of culture

C. THINKING AND PROBLEM-SOLVING (HIGH COG)
   (these characteristics do not typically have names and may not be measured in typical test situations)
   analysis
   synthesis
   evaluation
   decision-making

D. ATTITUDES AND VALUES (AFF)
   attitudes toward school
   attitudes toward self
   attitudes toward internal-external control
   vocational maturity
   interests
   prejudice
   tolerance of ambiguity
   value priority

E. LEARNING-RELATED BEHAVIOR (BEH)
   attendance
   time devoted to learning or problem-solving
   self-discipline
   initiative
   cooperativeness
   performance (in an actual or simulated setting; not paper-and-pencil)

Figure 1 Categories of Differentiated Outcomes and Examples Within Each

comprehension, (b) general knowledge and comprehension, (c) thinking and problem-solving, (d) attitudes and values, and (e) learning-related behavior. Figure 1 lists some possible outcomes in each category in order to make these categories clearer in meaning. (No attempt has been made to be exhaustive.)

It is recommended that at least one and preferably more than one variable in each category be included in every summative evaluation to at least broaden our knowledge of the differential effects of different educational interactions. The specific variables chosen for measurement should be based, insofar as possible, on the objectives of each of the treatments being compared. Unlike Ebel (1973), my recommendation is that program effects be measured in terms of more than just specific knowledge and comprehension goals and thus include objectives in all five areas. (It may be necessary first to get program designers to generate objectives in all five areas; otherwise some areas will continue to be typically overlooked.)

5. Some Current Evidence

It would be useful at this point to describe some of the work that has been completed using the notion of differentiated outcomes. While there may be a number of suitable examples in the literature, only a few will be described - all but one of which were done by my students and myself. That one was done by Worthen (1968) in an effort to compare two methods of task presentation in mathematics, the discovery approach versus the expository approach. I will not go into any detail regarding these inputs but will attend primarily to his differentiated outcomes. Five outcome variables were measured. These were (1) initial concept learning, (2) retention of concepts, (3) transfer of concepts, (4) transfer of heuristics (or strategies), and (5) attitudes toward content. The first variable

fits into the SPEC COG (see Figure 1) category, the fourth into the HIGH COG category, and the fifth into AFF. Variable 3 could be either SPEC COG or HIGH COG while variable 2 does not fit the categorization. Perhaps retention measures fit SPEC COG or perhaps they require a category of their own. Worthen found the two approaches to task presentation not to differ significantly on the initial learning (#1) or attitudes (#5) they produced; that the discovery group significantly exceeded the expository group on retention (#2) and transfer of heuristics (#4); and that the discovery group exceeded the expository group on transfer of concepts (#3) with the difference approaching significance. Overall, let us summarize by saying that differences were found mainly on HIGH COG variables as contrasted to SPEC COG or AFF variables.

Let me describe some of my studies so that a pattern may become detectable. Tuckman and Orefice (1973) compared four different procedures for teaching first year community college students a unit in accounting. These were: (a) self-instruction via tapes and booklets, (b) programmed instruction within a classroom setting, (c) programmed instruction and lectures within a classroom, and (d) traditional lecture - discussion instruction. As one proceeds from treatments (a) to (d), instruction is seen as being decreasingly student-initiated and increasingly teacher-initiated. Also, within this study, students were classified as either abstract or concrete in personality orientation. Measured outcomes included: (1) achievement of accounting objectives (SPEC COG), (2) student time spent in the instructional process (BEH), and (3) attitudes toward the form of instruction received (AFF). No differences between treatments, between personalities, or within their interaction was found on the achievement (SPEC COG) variable. On the attitude (AFF) measure, the self-study approach was found

to be clearly preferred and the lecture-programmed textbook approach liked least but even more striking was the interaction between treatment and personality. Abstract students disliked the lecture-programmed textbook most among the four treatments while concrete students preferred this approach; exactly the reverse was true for the programmed textbook without lecture approach: abstract students preferred it and concrete students disliked it. Consistent with their preference patterns, concrete students registered their dislike for the programmed approach alone by spending noticeably less time on it than on any of the other approaches.

Again, we find no differences on the SPEC COG variable with differences appearing this time on AFF and BEH measures. Moreover, our differentiated outcomes are sensitive not only to treatment differences but to personality differences in combination with treatments as well (what Cronbach and Gleser, 1965, call aptitude-treatment interactions and Hunt, 1971, calls matching models).

The missing element in the above study seemed to be a HIGH COG measure and so another study was done (Elliott and Tuckman, 1973) comparing individualized and traditional accounting instruction - this time in a naturalistically occurring way. A completely individualized program was compared to an entirely classroom-based one on a series of differentiated outcome measures including: (1) achievement on accounting objectives (SPEC COG), (2) student instructional time (BEH), (3) student attitudes toward school (AFF), and (4) a problem-solving measure that included measures of knowledge (SPEC COG), search initiative, time spent and solution length (BEHs), and adequacy of problem solutions (HIGH COG).

On the achievement measure (SPEC COG), again no treatment differences were found. No differences were found on the attitude (AFF) measure as

well. Individualized students were able to complete instruction in less time than classroom-taught students (the BEH measure).*

Let us focus our attention on the multi-faceted problem-solving measure. This was a take home problem in an aspect of accounting not covered in either treatment, viz., taxes on investments. Thus, it was unfamiliar in its specifics but within the range of subject-matter. It contained within it some unfamiliar terms. Students were asked to (1) provide definitions for the unfamiliar terms (SPEC COG), (2) indicate as many solutions as they could, in writing, for the problem (HIGH COG), and (3) keep track of and report time spent working on the problem (BEH). They were told to use whatever resources they chose to help them solve the problem but that they must (4) report the resources used (BEH). Finally, (5) problem-solution length was examined (BEH).

You can probably guess what happened. On the SPEC COG measure, no treatment differences were found. On the problem solutions (HIGH COG) as judged by accountants (with adequate scoring reliability demonstrated), individualized students did significantly better than classroom-taught students. Higher scores on each of the behavior measures were also obtained by individualized students.

So far we have reported on discovery and individualized instruction compared to more conventional approaches with the general finding being no differences in specific cognitive achievement between approaches but differences favoring one approach over the other in higher cognitive processes. Let us look closely at one more study. Tuckman, Cochran, & Travers (1974) compared open classroom instruction to conventional instruction at the elementary school level in terms of the following measures from students:

*Personality was also included in this study but produced no significant main effects or interactions.

(1) achievement in language arts, reading, and math (SPEC COG), (2) classification task skills (HIGH COG), (3) attitudes toward school (AFF), and (4) attitudes toward self (AFF). Significant differences were found only on the two affective measures - in both cases in favor of the open classroom group.

Although the above finding somewhat breaks the pattern, it can be seen as consistent with the treatments and hence in support of the differentiated outcome hypothesis. The open classroom program studied had just completed its first year and other findings in the study* showed that teacher social behavior was perhaps the most noticeable area of instructional change provoked by the short amount of in-service training received by teachers. Basic curriculum changes had not yet occurred. In turn, teacher behavior characterized more by warmth and acceptance resulted, not unreasonably, in affective gains among students without concomitant improvement in the higher cognitive area. Of course, by now the no-differences-in-achievement finding is also becoming a familiar outcome.

6. More to Come

Some other outcomes currently being (or recently having been) measured in our work include: student alienation (in modular-scheduled and conventional middle schools), attitudes about internal-external (fate) control (as the result of a high school political education program), teacher ratings of student behavior (in a junior high pre-vocational program for problem students), tolerance of ambiguity (in a high school senior year alternative program as contrasted to a conventional program), life skills - health, money management, and family life self-reported behaviors (among

*A number of measures of teacher behavior were also included in this study.

Puerto Rican and non Puerto Rican students), reasoning ability (among students in a school district-wide program using instruction based on behavioral objectives), and ability to complete an unfamiliar experiment (among students having access on a self-initiated basis or those not having this access to single-concept filmstrips). In each instance, in addition to the outcomes mentioned, others are chosen in an attempt to cover all five categories of the model.

7. The Moral

The differentiated outcome hypothesis is probably more a mandate or a proposition of good sense than a hypothesis. It is the function of a summative evaluation to demonstrate those outcomes that a program can achieve better than its alternatives in addition to any that it cannot achieve as well. It is then the province of those-who-provide or those-that-experience or those-that-finance the program to decide whether its outcomes are a sufficient basis for its continuation. However, it must be emphasized that all programs, even those that are considered alternatives to one another, cannot be expected to yield the same outcomes. One should be able to predict the kinds of outcomes that a program will maximize (if done optimally) by examining its goals and procedures. The job of the evaluator is not to decide which outcomes are the most important but to demonstrate which outcomes, if any, are maximized. While those who theorize about learning and instruction typically concentrate on educational inputs, it is left to evaluators to be concerned about outcomes.* And since the list of possible outcome measures is endless, the evaluator must have a basis for choosing the ones to examine. Instead of accepting conventional achievement test scores as the "shoe" and maintaining "if the shoe fits, wear it," evaluators must be willing and able to examine the full range of "sizes" and "styles."

*Although as Lohnes (1973 suggests, evaluators must also determine that the inputs or processes are consistent with the name that has been used to label them.

References

Cronbach, L. J. Test validation, In R. L. Thorndike (Ed.), Educational Measurement: second edition. Washington, D.C.: American Council on Education, 1971, pp. 443-507.

Cronbach, L. J. & Gleser, G. C. Psychological tests and personnel decisions (2nd ed.). Urbana: University of Illinois Press, 1965.

Ebel, R. L. The future of the measurement of abilities II. Educational Researcher, 1973, 2 (No. 3), 5-12.

Elliot, J. M. and Tuckman, B. W. Differentiated outcomes resulting from individualized instruction at a two-year college. Unpublished manuscript, 1973.

Gagné, R. M. Instruction based on research in learning. Engineering Education, 1971 (March), 520 ff.

Gagné, R. M. Domains of learning. Interchange, 1972, 3, 3-13.

Glaser, R. Individuals and learning: The new aptitudes. Educational Researcher, 1972, 1, (No. 6), 5-13.

Hunt, D. E. Matching models in education. Toronto: Ontario Institute for Studies in Education, 1971.

Lohnes, P. R. Evaluating the schooling of intelligence. Educational Researcher, 1973, 2 (No. 2), 6-11.

Tuckman, B. W. and Orefice, D. S. Personality structure, instructional outcomes, and instructional preferences. Interchange, 1973, 4, 43-48.

Tuckman, B. W., Cochran, D. W. and Travers, E. J. Evaluating the open classroom. Journal of Research and Development in Education, 1974 (in press).

Tyler, R. W. The functions of measurement in improving instruction. In E. F. Lindquist (Ed.), Educational measurement. Washington, D.C.: American Council on Education, 1951, pp. 47-67.

Worthen, B. Discovery and expository task presentation in elementary mathematics. Journal of Educational Psychology Monograph Supplement, 1968 (part 2), 59, 1-13.